Estimation of linkage disequilibrium in a sample of the United Kingdom dairy cattle population using unphased genotypes

Estimation of linkage disequilibrium in a sample of the United Kingdom dairy cattle population using unphased genotypes

A. Tenesa, S. A. Knott, D. Ward, D. Smith, J. L. Williams and P. M. Visscher

J ANIM SCI 2003, 81:617-623.

Estimation of linkage disequilibrium in a sample of the United Kingdom dairy cattle population using unphased genotypes[1]

A. Tenesa*[2], S. A. Knott*, D. Ward†, D. Smith†, J. L. Williams†, and P. M. Visscher*

* Institute of Cell, Animal and Population Biology, University of Edinburgh, EH9 3JT, Scotland, U.K. and †Roslin Institute, Midlothian EH25 9PS, Scotland, U.K.

ABSTRACT: The association between genetic marker alleles was estimated for two regions of the bovine genome from a random sample of 50 young dairy bulls born in the United Kingdom between 1988 and 1995. Microsatellite marker genotypes were obtained for six markers on chromosome 2 and seven markers on chromosome 6, spanning 38 and 20 cM, respectively. Two different methods, which do not require family information, were used to estimate population haplotype frequencies. Haplotype frequencies were estimated for pairs of loci using the expectation-maximization algorithm and for all linked loci using a Bayesian approach via a Markov chain-Monte Carlo algorithm. Significant ($P = 0.0007$) linkage disequilibrium was detected between pairs of loci in syntenic groups (that is, loci in the same linkage group), extending to about 10 cM. No significant linkage disequilibrium was detected between markers in nonsyntenic regions. Given the observed level of linkage disequilibrium, mapping methods based on population-wide association might provide a better resolution than traditional quantitative trait loci mapping methods in the U.K. dairy cattle population and may reduce the required sample sizes of the experiments.

Key Words: Dairy Cattle, Linkage Disequilibrium, Mapping, Markers

 J. Anim. Sci. 2003. 81:617–623

Introduction

Linkage disequilibrium (**LD**) mapping methods use LD at the population level to map trait loci. These methods have higher power (Risch and Merikangas, 1996) and higher resolution than traditional linkage methods because they use information based on a larger number of meioses. The power of LD mapping methods depends on population parameters such as allele frequencies at the marker and trait loci and level of LD. The achievable resolution depends on the extent of disequilibrium between marker and trait loci.

Although the extent and patterns of LD have been extensively studied in human populations, (Daly et al., 2001; Jeffreys et al., 2001) farm animal populations have been rarely studied.

Farnir et al. (2000) and McRae et al. (2002) studied the extent of LD in the Dutch black-and-white dairy cattle population and in two sheep populations, respectively. Both these studies used family information to infer the most likely phase of the dams. However, family information is not always available and, if available, collecting the additional family members required may be an inefficient use of resources.

In this study, we estimate the extent of LD in the U.K. dairy cattle population. This will determine the feasibility of LD mapping methods in this population and the marker density required for LD mapping to be effective. We illustrate the use of statistical methods that do not require family information to infer population haplotype frequencies as an alternative to family-based haplotyping methods. These methods to estimate haplotype frequencies are relatively efficient compared to those that require family information (Hill, 1974; McKeigue, 2001). We applied these methods in a small data set and assessed the extent of LD in two regions of the genome of 50 randomly selected dairy cattle bulls that were being progeny tested. They were assumed to produce a representative sample of the future extent of LD in the U.K. dairy cattle population.

Materials and Methods

Data

Data comprised genotypes from 50 Holstein bulls that were being progeny tested. The bulls were born between

[1]We are grateful for support from the Medical Research Council Human Genetic Unit, the Biotechnology and Biological Sciences Research Council, and the Royal Society. We thank S. Brotherstone for help with the pedigree data and A. Carothers for comments on the manuscript.

[2]Correspondence: West Mains Rd. (E-mail: albert.tenesa@ed.ac.uk).

Received May 23, 2002.
Accepted October 28, 2002.

Table 1. Genetic map, number of alleles at the marker locus, percentage of missing values, observed heterozygosity at the marker loci, expected heterozygosity under Hardy-Weinberg equilibrium (HWE) and significance level (P) of the test for departures from HWE for chromosome 2

Marker	TGLA226	BMS829	BMS2519	BM2113	IDVGA37	IDVGA2
Genetic map, cM	80	91.5	101.5	106.2	108.2	117.8
Number of alleles	5	5	5	6	3	5
Missing values, %	28	28	34	26	18	32
Observed heterozygosity	0.61	0.33	0.58	0.81	0.39	0.59
Expected heterozygosity	0.79	0.40	0.70	0.76	0.39	0.72
Departures from HWE, *P*	<0.001	0.08	<0.001	0.57	0.75	0.45

1988 and 1995. Bulls were genotyped at six marker loci on chromosome 2 and at seven marker loci on chromosome 6. Genotyping was carried out as described by Wiener et al. (2000), and marker identities are given in Tables 1 and 2. Each bull pedigree was known up to three generations. Grandparents were assumed unrelated. Relationships between bulls are shown in Tables 3 and 4. Genetic distances (Kosambi map function) between markers were obtained from the map MARC97 (Kappes et al., 1997).

Haplotype Frequency Estimation and Hardy-Weinberg Equilibrium Proportions

Maximum likelihood estimates of all 78 ($13 \times [13 - 1]/2$) two-marker loci haplotype frequencies were estimated by employing the expectation-maximization (**EM**) algorithm (Excoffier and Slatkin, 1995) as implemented in Gold (Abecasis and Cookson, 2000). Relationships between bulls were ignored when estimating haplotype frequencies. We tried to obtain maximum likelihood estimates of six-loci and seven-loci haplotype frequencies for chromosomes 2 and 6, respectively, using Arlequin (Genetics and Biometry Lab, University of Geneva, Switzerland). The algorithm failed to reach a global maximum likelihood estimate of the haplotype frequencies; therefore, estimates were not used in this study. We did not try to estimate fewer than six- and seven-loci haplotype frequencies other than two-loci haplotype frequencies.

Bayesian estimates of six- and seven-loci haplotype frequencies for chromosome 2 and 6, respectively, were obtained using PHASE (Stephens et al., 2001). No attempt to estimate LD among nonsyntenic loci (loci in a different linkage group) using the Bayesian approach was made. Haplotypes were reconstructed 10 independent times to ensure that the results obtained were robust even if the algorithm was not converging, as suggested by Stephens et al. (2001). We ran the algorithm for 10^7 iterations after a burn-in period of 10^4 and kept estimates from every 100th iteration. The program PHASE assumes, by default, a stepwise mutation model; however, this assumption was relaxed by using a parent-independent mutation model in which each microsatellite allele has the same chance to mutate to any of the other alleles. Although a stepwise mutation model is more appropriate for microsatellite markers if the length of each microsatellite allele is known, we did not know the actual length of the microsatellite alleles in these data; therefore, this model could not be assumed.

Departures from Hardy-Weinberg equilibrium (**HWE**) proportions were tested using an exact test as described by Guo and Thompson (1992). This algorithm is implemented in Arlequin (Genetics and Biometry Lab, University of Geneva). The Hardy-Weinberg Equilibrium is an assumption of the EM algorithm, and departures from HWE might lead to biased estimates of haplotype frequencies (Excoffier and Slatkin, 1995). In addition, departures from HWE can be an indication of population stratification, selection of the locus or linked locus, different fertility of parents or different allele frequencies in male and female parents, finite population size, and so on.

Table 2. Genetic map, number of alleles at the marker locus, percentage of missing values, observed heterozygosity at the marker loci, expected heterozygosity under Hardy-Weinberg equilibrium (HWE) and significance level (*P*) of the test for departures from HWE for chromosome 6

Marker	RM28	BM415	CSN3	BM1236	BMS511	AFR227	BM8124
Genetic map, cM	74.3	76.3	82.6	83.9	89.8	90.4	94.2 cM
Number of alleles	4	7	3	4	5	6	2
Missing values, %	18	4	8	14	10	6	0
Observed heterozygosity	0.66	0.67	0.35	0.60	0.78	0.34	0.16
Expected heterozygosity	0.67	0.79	0.40	0.57	0.74	0.74	0.17
Departures from HWE, *P*	0.61	<0.001	0.26	0.31	0.81	<0.001	0.99

Table 3. Number (N) of maternal-grand-sire and half-sib groups in the sample

n	N of maternal-grand-sire groups with n bulls	N of paternal half-sib groups with n bulls
1	18	23
2	5	6
3	3	3
6	1	1
7	1	0
Total	28	33

Level of Linkage Disequilibrium

Hedrick's normalized measure of disequilibrium (Hedrick, 1987) was obtained from the estimates of the two-loci haplotype frequencies. Hedrick's normalized measure of disequilibrium is the extension to multiallelic loci of the normalized measure of disequilibrium defined by Lewontin (1964). It is defined as follows:

$$D' = \sum_{m=1}^{k} \sum_{n=1}^{l} m_m q_n \, |D'_{mn}| \qquad [1]$$

where k and l are the number of alleles at locus M and Q, respectively, m_m *and* q_n are the population allele frequencies of allele m at locus M and allele n at locus Q, respectively. $|D'_{mn}|$ is the absolute value of Lewontin's normalized measure:

$$D'_{mn} = \frac{D_{mn}}{D_{mn}^{\max}} = \frac{(h_{mn} - m_m q_n)}{D_{mn}^{\max}} \qquad [2]$$

where h_{mn} is the estimated population frequency of the haplotype $M_m Q_n$, and $D^{max}{}_{mn}$ is the maximum amount of disequilibrium possible between allele m at locus M and allele n at locus Q that equals:

$$\mathrm{D}_{\mathrm{mn}}^{\max} = \begin{bmatrix} \min\{m_m q_n, (1 - m_m)(1 - q_n)\}; \, D_{mn} < 0 \\ \min\{m_m(1 - q_n), (1 - m_m) q_n\}; \, D_{mn} > 0 \end{bmatrix} \qquad [3]$$

Table 4. Additive genetic relationships among bulls calculated using the three-generation pedigree

Additive genetic relationships	Number of relationships with additive genetic relationships
0.00000	837
0.01563	15
0.03130	70
0.06250	135
0.07813	2
0.09380	6
0.12500	105
0.15630	8
0.18750	7
0.25000	31
0.31250	3
0.50000	6

To test the statistical significance of the allelic association, we compared the statistic $S = 2ln(L_{LD}/L_{LE})$ to a χ^2 distribution with $(k - 1) \times (l - 1)$ degrees of freedom (Slatkin and Excoffier, 1996). Assuming random mating, L_{LD} is the likelihood computed using the haplotype frequencies found by the EM algorithm, and L_{LE} is the likelihood under the assumption of linkage equilibrium. We assumed that the available sample size was large enough for asymptotic assumptions to hold.

We performed a large number of tests (n = 78); therefore, we applied a Bonferroni correction to obtain an appropriate significance level for association between each pair of marker loci. The individual test significance level after correction to give a total significance level (γ) of 0.05 was $P = 1 - (1 - \gamma)^{1/n} = 0.0007$, where n was the total number of tests performed. Because some tests are likely to be correlated, our stringent threshold is expected to be conservative with respect to the type-I error rate.

Results

Departures from Hardy-Weinberg Equilibrium

Thirteen microsatellite markers spanning bovine chromosomes 2 and 6 were genotyped on 50 dairy bulls. Genetic positions of the markers, number of alleles at each locus, percentage of missing values, observed heterozygosities, expected heterozygosities under HWE for the observed population allele frequencies, and significance level of the test for departures from HWE proportions are shown in Tables 1 and 2. The 13 markers had an average observed heterozygosity of 0.53 and an average expected heterozygosity of 0.60. The average distance between markers was 4.4 cM across a length of 57.7 cM. The mean number of alleles was 4.6.

Nine of the 13 markers studied showed a deficiency of heterozygotes; however, only four of these nine showed significant ($P < 0.001$) departures from HWE proportions. Relatedness between individuals in our sample and the small effective population size of the worldwide dairy cattle population could be the cause of the observed deficiency of heterozygotes.

Linkage Disequilibrium Between Syntenic Marker Loci Using the EM Algorithm

Figure 1 shows a plot of the extent of disequilibrium (D′) vs genetic map distance measured in cM (genetic map distance is hereafter referred to as genetic distance). The average D′ was 44%. The most remarkable observation was that D′ did not seem to vary as a function of the genetic distance. We fitted a nonlinear equation of type $y = a + be^{-cx}$ using nonlinear regression as implemented by Genstat's FITCURVE directive (Genstat 5 Committee, 1993), where y is D′ and x is genetic distance in cM. Note that y tends to a when x tends to infinity and y tends to $a + b$ when x tends to zero. Only a was ($P < 0.0001$) different from zero. The estimated

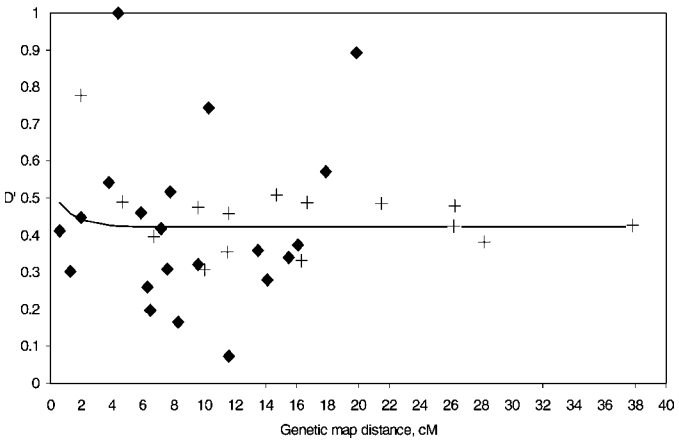

Figure 1. Relationship between genetic distance (cM) and level of linkage disequilibrium (D′). The plotted line represents the fitted line. Crosses and diamonds represent comparisons between pairs of loci on chromosome 2 and chromosome 6, respectively.

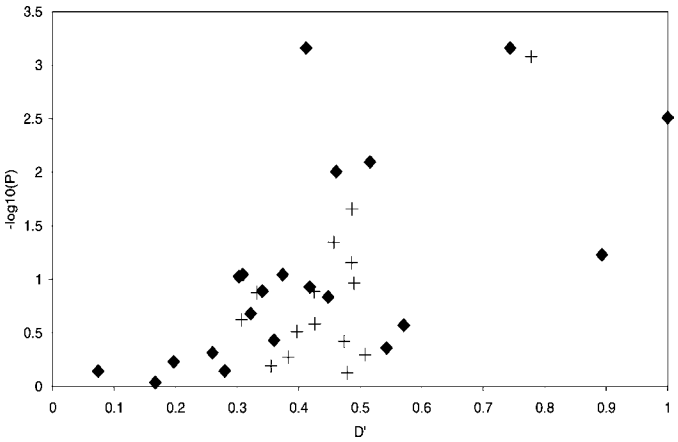

Figure 3. Relationship between level of significance ($-\log_{10}[P]$) and level of linkage disequilibrium (D′) for syntenic loci pairs. Crosses and diamonds represent comparisons between pairs of loci on chromosome 2 and chromosome 6, respectively.

parameter values are 0.42 ± 0.06 for a, 0.11 ± 0.18 for b, and 0.76 ± 0.59 for e^{-c}. The fit of $y = a$ and $y = a + be^{-cx}$ was compared using a likelihood ratio test. The fit of the two curves was not significantly different.

The level of association ($-\log_{10}[P]$) showed a clearer correlation with distance (Figure 2) but still highly variable, especially for the smallest distances. All $P < 0.01$ ($-\log_{10}[P] > 2$ in Figure 2) correspond to genetic distances smaller than 10.3 cM. Only two pairs of markers were in significant linkage disequilibrium after accounting for multiple testing. These were BM1236-BM8124 ($P = 0.0007$; intermarker distance = 10.3 cM) and BMS511-AFR227 ($P = 0.0007$; intermarker distance = 0.6 cM) on chromosome 6. Before correcting for multiple testing, there was a total of eight pairs in significant association at the 5% level. Three of these pairs were on chromosome 2 and five on chromosome 6.

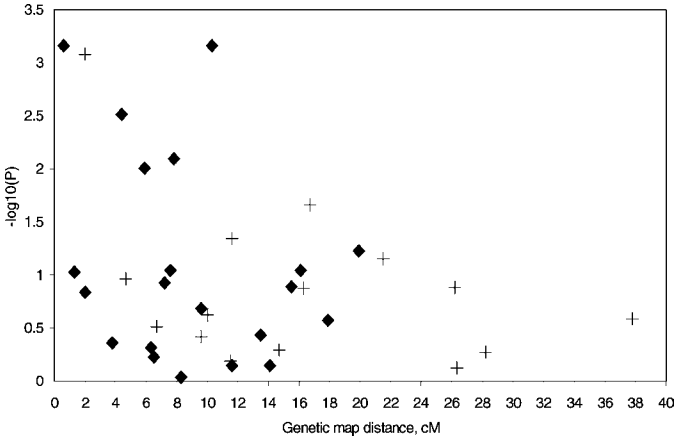

Figure 2. Relationship between level of significance ($-\log_{10}[P]$) and genetic distance (cM) for syntenic loci pairs. Crosses and diamonds represent comparisons between pairs of loci on chromosome 2 and chromosome 6, respectively.

Although we observed a high average level of disequilibrium, only two pairs of loci showed a significant association. To test whether the mean level of disequilibrium observed was significant, we calculated: 1) the sum of the 36 statistics ($6 \times 7/2$ and $5 \times 6/2$ from chromosome 6 and 2, respectively; ($X^2 = 646$) and 2) the sum of the 36 associated degrees of freedom (**df** = 456). This overall test for average level of LD across all pairs of syntenic loci was highly significant ($P[\chi^2_{456\ df} \geq (X^2 = 646] << 10^{-7}$), indicating that the mean level of disequilibrium was different from zero and that we lacked power when testing individual pairs.

Figure 3 shows a plot of $-\log_{10}(P)$ for each pair of marker loci as a function of D′. Significant LD tended to increase with D′, although it was very variable. This variance seemed to depend on the value of D′. Pairs of loci with larger values of D′ showed more variable levels of significance.

Linkage Disequilibrium Between Syntenic Marker Loci Using the Bayesian Algorithm

Figures 4 and 5 show the comparison in the estimates of D′ for chromosome 2 and 6, respectively, using the maximum likelihood and Bayesian approach to estimate haplotype frequencies. Maximum-likelihood estimates are plotted as single points, and Bayesian estimates are plotted as the mean of D′ obtained from 10 independent estimates of the haplotype frequencies with lines indicating two standard deviations. Since we have only one estimate of D′ when using the EM algorithm, formal comparisons between both estimates cannot be performed. However, qualitative comparisons can be done and the general picture is the same regardless of the estimation method used.

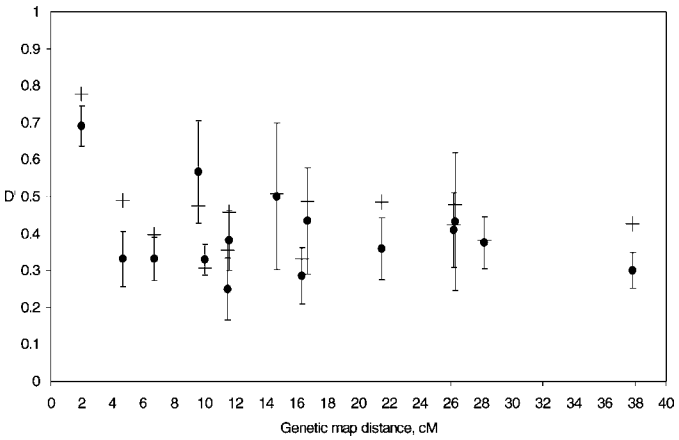

Figure 4. Comparison of the estimates of D′ obtained when using population haplotype frequencies estimated by the maximum-likelihood (crosses) and Bayesian approaches (circles) for chromosome 2. Each circle is the mean of 10 runs of the program PHASE and the lines are ± 2 SD.

Another important observation is that the variance of D′ is highly variable for chromosome 2, but not for chromosome 6 (note that some of the estimates have variance equal to zero). This probably reflects more missing values for chromosome 2 than for chromosome 6 (Tables 1 and 2).

Results using a stepwise mutation model (results not shown) were not significantly different from those from the parent-independent mutation model. This suggests that the algorithm is relatively insensitive to the underlying assumptions about the mutation model.

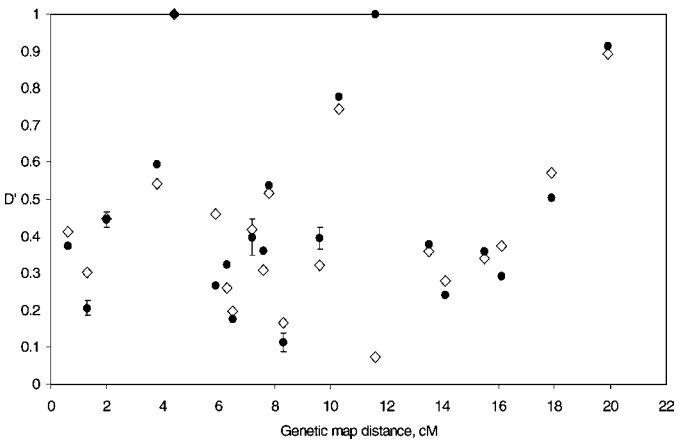

Figure 5. Comparison of the estimates of D′ obtained when using population haplotype frequencies estimated by the maximum-likelihood (diamonds) and Bayesian approaches (circles) for chromosome 6. Each circle is the mean of 10 runs of the program PHASE and the lines are ± 2 SD.

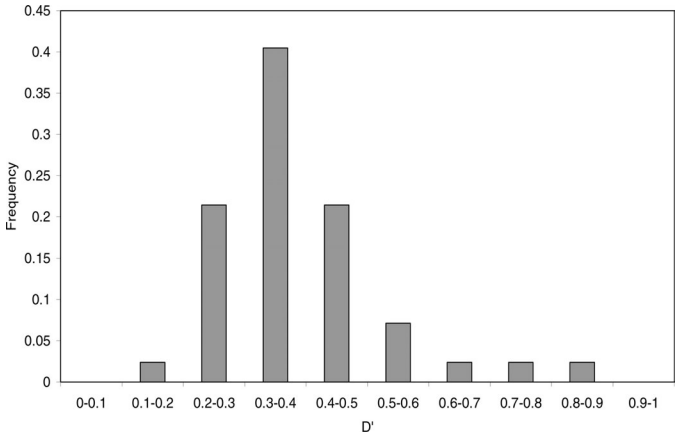

Figure 6. Distribution of D′ values observed between pairs of nonsyntenic loci.

Linkage Disequilibrium Between Nonsyntenic Marker Loci Using the Expectation-Maximization Algorithm

Figure 6 shows the distribution of D′ values observed between pairs of nonsyntenic loci. We estimated the mean level of LD between nonsyntenic loci, measured as D′, to be 39%. None of the loci pairs showed significant association between alleles. Indeed, the most significant association was for the pair BM2113-BM1236 ($P = 0.03$; D′ = 0.53). The sum of the 42 statistics obtained between nonsyntenic loci was 548, and the sum of the 42 associated df was 539. The overall level of association between pairs of nonsyntenic loci was not significant ($P[\chi^2_{539\ df} \geq (X^2 = 548] = 0.39$). In addition to this overall test, we performed a Fisher's combined probability test (Fisher, 1970) for syntenic and nonsyntenic groups that gave similar results (results not shown). Overall, average levels of LD were fairly similar between syntenic and nonsyntenic loci; however, association could be statistically detected between syntenic loci, but not between nonsyntenic loci, even when the D′ values were similar.

Discussion

Our results show that LD mapping methods could be successfully applied to future U.K. dairy cattle populations with the available density of microsatellite markers. Significant linkage disequilibrium was found only for genetic distances smaller than about 10 cM, and significant association was never found between nonsyntenic loci. This would have important implications for LD mapping. Firstly, the mapping resolution achievable with this level of disequilibrium would be finer than with traditional QTL-mapping methods. Secondly, if the lack of significant association found here between loci on chromosomes 2 and 6 was the same across the whole genome, then the number of false positives due to allelic associations between unlinked loci would be small when applying LD methods to map trait loci.

Some aspects of our results differ from those reported by Farnir et al. (2000). First, they found extensive significant LD between both syntenic and nonsyntenic loci. Second, they found average D′ values in the same range as ours only for genetic distances <5 cM. Third, they found that only those D′ values for the more distant syntenic markers were similar to those between nonsyntenic markers. These differences might arise because of two reasons. First, our sample is more related than theirs, and therefore showed larger identical by descent regions. They used two different samples for estimating the extent of LD. One sample was composed of bull-dams and the other of cows selected from the general population. Although their first data set might have a level of relatedness as high as that in our data, it is unlikely that cows in their second data set were as related as our bulls. Relatedness between individuals can cause an increase in the level of LD, even between unlinked loci, because larger portions of the genome are identical between related individuals. Second, the sample size of both studies is very different and a comparison might be difficult and even inappropriate. The expectation of D' under equilibrium is zero; however, its sampling variance depends on the sample size from which it is estimated: The larger the sample size, the smaller the sampling variance. If the sampling variance is large, then it is more likely that, just by chance, the estimated value for D′ differs from zero. Weir and Hill (1980) derived the variance of R, the correlation of gene frequencies, for biallelic loci. Their arguments about the two sampling processes involved in estimating LD can be extended to a different measure of disequilibrium, say D′. For closely linked loci, the variance of R is approximately $1/(1 + 4N_ec) + 1/n$, where N_e is the effective population size, c is the recombination fraction between the two loci, and n is the sample size. The variance of R is due to two different sampling processes, one that reflects the finite size of the population $(1/[1 + 4N_ec])$ and another that reflects that a limited sample of the population $(1/n)$ has been drawn (from which disequilibrium and allele frequencies have been estimated). It is worth noting that n is either a sample of n identified chromosomes or n unphased individuals from which disequilibrium and allele frequencies have been estimated. Additionally, for D′, the difference from its expected value under equilibrium is aggravated by the fact that D′ uses the absolute value of D'_{mn}. Even small deviations from equilibrium between pairs of alleles accumulate, leading to an upwards bias in the estimate of D′.

We believe that lack of statistical power, especially after correcting for multiple testing, and an upwards bias (due to the small sample size) in the estimate of D′ is the reason why the larger D′ values observed did not correspond to more significant allelic associations. We assumed that all the tests performed were independent; however, tests between loci on the same chromosome are correlated, especially if the distance between loci is not large as in our data. The significance thresholds we applied after correction are, therefore, very conservative as the number of independent tests actually performed was smaller than assumed.

It is unlikely that the departures from HWE expectations we observed led to an important degree of bias in the estimates of haplotype frequencies. The only problem when estimating haplotype frequencies from genotypes comes from individuals that are heterozygous at the loci considered. In this situation, haplotype frequencies cannot be directly counted because it is not possible to distinguish between the two different diplotypes (i.e., an individual with the two-loci genotype AaBb could have diplotype Ab/aB or AB/ab). In this case, the EM algorithm iteratively estimates the frequencies of the different haplotypes until the likelihood of the data is maximized and, therefore, maximum likelihood haplotype frequencies are obtained. When there is an excess of homozygotes, the number of doubly heterozygous individuals to be resolved is smaller. Consequently, there is little or no bias in the haplotype frequency estimates caused by deviations from HWE due to an excess in homozygosity (Osier et al., 1999; Fallin and Schork, 2000).

Six- and seven-loci maximum likelihood haplotype frequencies for chromosome 2 and 6, respectively, could not be obtained. This was because the algorithm failed to reach a global maximum. After each step of the EM algorithm, the likelihood of the data increases (Dempster et al., 1977); however, if the likelihood surface is concave or very flat, then there is no guarantee that a global maximum is reached. Generally, there is no obvious way of knowing whether the estimated maximum is just a local or a global maximum. In order to be sure that a global maximum is reached, the algorithm is usually started several times from different starting points, and the solution with the maximum likelihood is assumed to be the global maximum. In our case, although the likelihood of the data was the same for different runs, we obtained different haplotype frequencies in each of the runs. This suggests that the likelihood surface was very flat due to the insufficient amount of data or dependencies between the data, and that the iterative process stopped before reaching the global maximum.

Differences observed between the maximum likelihood and Bayesian approaches were small and the general conclusions obtained from both estimation procedures were essentially the same. Differences observed between both approaches are slightly larger for chromosome 2, which has more missing values, than for chromosome 6. This might suggest that the amount of data for some loci on chromosome 2 is too small and this is reflected in the slightly larger discrepancies between both approaches. An advantage of the Bayesian approach is that it provides estimates of the uncertainty associated with each phase, at the cost of a much larger computing time. An advantage of the maximum likelihood over the Bayesian approach is that implementation of the testing procedure is straightforward in the

maximum likelihood framework. Therefore, the decision about the most appropriate method would depend on the intended use of the haplotype frequencies. For example, if one just wanted to test for the presence of LD, then the maximum likelihood approach seems adequate and straightforward, but if one wanted to compare haplotype frequencies in a cases/control design then an estimate of the uncertainty of each phase would be necessary.

The fact that the disequilibrium parameter (D′) did not depend on distance (cM) but *P* did depend on distance (Figures 1 and 2), and that similar values of D′ were observed between syntenic and nonsyntenic loci (but significance level was different), suggests that the utility of D′ to assess the amount of disequilibrium is limited. This is important if assessment of disequilibrium is done as a preliminary study to determine, for example, the marker density required for a mapping study. In this case, the correlation between *P* and distance will give a clearer "picture" of the marker density required.

The region of chromosome 6 where we detected the most significant LD has been reported to harbor QTL influencing milk, fat, and protein yield in the U.K. dairy population (Wiener et al., 2000) and other populations, such as the Israeli Holstein population (Ron et al., 2001). This suggests that selection for milk production traits could have generated LD in this region, which was detectable even with the large amount of background LD observed.

Implications

Fine mapping of trait loci in outbred populations relies on population-based samples for which linkage disequilibrium between trait and marker loci is expected to occur at smaller distances than in family-based samples. The amount of linkage disequilibrium between marker loci in a population provides information about the marker density required to perform the mapping study. In livestock populations, this type of study has always been done using family information to infer phase; however, this procedure requires typing additional family members. Even if possible, typing extra members might be an inefficient use of resources, especially when statistical methods such as those described in this study are known to perform reasonably well.

Literature Cited

Abecasis, G. R., and W. O. C. Cookson. 2000. Gold—Graphical overview of linkage disequilibrium. Bioinformatics 16:182–183.

Daly, M. J., J. D. Rioux, S. E. Schaffner, T. J. Hudson, and E. S. Lander. 2001. High-resolution haplotype structure in the human genome. Nat. Genet. 29:229–232.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. B. 39:1–38.

Excoffier, L., and M. Slatkin. 1995. Maximum-likelihood-estimation of molecular haplotype frequencies in a diploid population. Mol. Biol. Evol. 12:921–927.

Fallin, D., and N. J. Schork. 2000. Accuracy of haplotype frequency estimation for biallelic loci, via the expectation-maximization algorithm for unphased diploid genotype data. Am. J. Hum. Genet. 67:947–959.

Farnir, F., W. Coppieters, J. J. Arranz, P. Berzi, N. Cambisano, B. Grisart, L. Karim, F. Marcq, L. Moreau, M. Mni, C. Nezer, P. Simon, P. Vanmanshoven, D. Wagenaar, and M. Georges. 2000. Extensive genome-wide linkage disequilibrium in cattle. Genome Res. 10:220–227.

Fisher, R. A. 1970. Statistical Methods for Research Workers. 14th ed. Oliver and Boyd, Edinburgh, U.K.

Genstat 5 Committee. 1993. Genstat 5 Reference Manual. Clarendon Press, Oxford, U.K.

Guo, S. W., and E. A. Thompson. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48:361–372.

Hedrick, P. W. 1987. Gametic disequilibrium measures—proceed with caution. Genetics 117:331–341.

Hill, W. G. 1974. Estimation of linkage disequilibrium in randomly mating populations. Heredity 33:229–239.

Jeffreys, A. J., L. Kauppi, and R. Neumann. 2001. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. Nat. Genet. 29:217–222.

Kappes, S. M., J. W. Keele, R. T. Stone, T. S. Sonstegard, T. P. L. Smith, R. A. McGraw, N. L. LopezCorrales, and C. W. Beattie. 1997. A second-generation linkage map of the bovine genome. Available: http://www.ri.bbsrc.ac.uk/cgi-bin/mapviewer?species=cattle. Accessed Oct. 12, 2001.

Lewontin, R. C. 1964. The interaction of selection and linkage. I. General considerations; heterotic models. Genetics 49:49–67.

McKeigue, P. M. 2001. Efficiency of estimation of haplotype frequencies: Use of marker phenotypes of unrelated individuals versus counting of phase-known gametes. Am. J. Hum. Genet. 67:1626–1627.

McRae, A. F., J. C. McEwan, K. G. Dodds, T. Wilson, A. M. Crawford, and J. Slate. 2002. Linkage disequilibrium in domestic sheep. Genetics 160:1113–1122.

Osier, M., A. J. Pakstis, J. R. Kidd, J. F. Lee, S. J. Yin, H. C. Ko, H. J. Edenberg, R. B. Lu, and K. K. Kidd. 1999. Linkage disequilibrium at the Adh2 and Adh3 loci and risk of alcoholism. Am. J. Hum. Genet. 64:1147–1157.

Risch, N., and K. Merikangas. 1996. The future of genetic studies of complex human diseases. Science 273:1516–1517.

Ron, M., D. Kliger, E. Feldmesser, E. Seroussi, E. Ezra, and J. I. Weller. 2001. Multiple quantitative trait locus analysis of bovine chromosome 6 in the Israeli Holstein population by a daughter design. Genetics 159:727–735.

Slatkin, M., and L. Excoffier. 1996. Testing for linkage disequilibrium in genotypic data using the expectation-maximization algorithm. Heredity 76:377–383.

Stephens, M., N. J. Smith, and P. Donnelly. 2001. A new method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68:978–989.

Weir, B. S., and W. G. Hill. 1980. Effect of mating structure on variation in linkage disequilibrium. Genetics 95:477–488.

Wiener, P., I. Maclean, J. L. Williams, and J. A. Woolliams. 2000. Testing for the presence of previously identified QTL for milk production traits in new populations. Anim. Genetics 31:385–395.